ISBN 82-553-0647-1 nr 1
Stat.Res.Report Feb. 1988

CONFIDENCE INTERVALS AND CONFIDENCE BANDS FOR THE SURVIVAL CURVE BASED ON TRANSFORMATIONS

by

Ørnulf Borgan[1] and Knut Liestøl[2]

1) Department of Mathematics, University of Oslo
2) Department of Informatics, University of Oslo

STATISTICAL RESEARCH REPORT - Matematisk institutt,
Universitetet i Oslo

Abstract

Pointwise confidence intervals and simultaneous confidence bands for the survival distribution based on transformations are considered, and their small sample performance is compared with that of their non-transformed counterparts. For the confidence intervals and "the equal precision" bands a substantial improvement is obtained by using one of the transformed versions. For bands of the Hall-Wellner type there seems to be less reason for using transformations.

Keywords: Censored survival data; confidence intervals; confidence bands; Kaplan-Meier estimator; small sample properties; survival distribution.

1. Introduction

For three decades, Kaplan and Meier's (1958) estimator for the survival distribution has been a cornerstone in the statistical analysis of censored survival data. A rigourous study of the asymptotic distribution of this estimator was first undertaken by Breslow and Crowley (1974) for the particular case of random censorship, and later by Aalen and Johansen (1978) and Gill (1980) for quite general censoring patterns.

Based on these asymptotic distributional results, a number of authors have derived and studied pointwise confidence intervals and simultaneous confidence bands for the survival function. Thomas and Grunkemeier (1975) derived pointwise confidence intervals and studied their small sample properties by Monte Carlo simulations. Suggestions for simultaneous confidence bands for the survival curve have been put forward by Gillespie and Fisher (1979), Hall and Wellner (1980) and Nair (1981, 1984). Csörgő and Horváth (1986) introduced a class of bands which shows the relationship between most of these proposals, and they suggested some modified versions of the confidence bands. An extensive study of the small sample properties of some of the most used confidence bands was carried out by Nair (1984).

In a footnote to their paper, Thomas and Grunkemeier (1975) suggested the use of the arcsine-square root transformation to improve the small sample properties of pointwise confidence intervals for the survival probabilities. Later Nair (1984) proposed the use of the same transformation in connection with simultaneous confidence bands. Kalbfleisch and Prentice (1980), on the other hand, suggested the use of the log-minus-log transformation for pointwise confidence intervals. It is the purpose of the present paper to study the small sample properties of confidence intervals and confidence bands for the survival function based on such transformations. For the confidence bands we restrict our attention to bands of the Hall-Wellner type and to the "equal precision bands" (Nair, 1984) which are proportional to the pointwise intervals.

2. Pointwise confidence intervals and simultaneous confidence bands: asymptotic results.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed positive random variables (lifetimes) with absolutely continuous distribution function F, survival function $S = 1-F$ and hazard rate function $\alpha = F'/S$. We consider the set-up with right censoring, where X_i is only observed exactly if it does not exceed a (possibly random) censoring time Z_i. Thus our data are $(\tilde{X}_i, D_i)$; $i=1,2,\ldots,n$; where $\tilde{X}_i = X_i \wedge Z_i$ and $D_i = I\{\tilde{X}_i = X_i\}$. Here $s \wedge t = \min\{s,t\}$ and $I\{\cdot\}$ the indicator function. Let

$$Y_n(t) = \sum_{i=1}^{n} I\{\tilde{X}_i > t\} \tag{2.1}$$

denote the number at risk at $t-$. Then

$$\hat{S}_n(t) = \prod_{\{i:\tilde{X}_i < t, D_i = 1\}} \left(1 - \frac{1}{Y_n(\tilde{X}_i)}\right) \tag{2.2}$$

is the Kaplan-Meier estimator. Its variance is estimated by the Greenwood formula

$$\hat{\tau}_n^2(t) = [\hat{S}_n(t)]^2 \, \hat{\sigma}_n^2(t) ,$$

where

$$\hat{\sigma}_n^2(t) = \sum_{\{i:\tilde{X}_i < t, D_i = 1\}} \{Y_n(\tilde{X}_i)[Y_n(\tilde{X}_i) - 1]\}^{-1} . \tag{2.3}$$

The censoring is assumed to be independent in the sense of Kalbfleisch & Prentice (1980, p. 120), see also Gill (1980, Theorem 3.1.1). This is the case for all the usual types of right censoring, like random censorship and (progressive) censoring of Type I and II (Gill, 1980, Corollary 3.1.1).

We will briefly review the asymptotic properties of the Kaplan-Meier estimator. Assume that there exists a constant T and a function y with $y(T)>0$ such that $Y_n(t)$ defined in (2.1) satisfies

$$\sup_{t\in[0,T]} |Y_n(t)/n - y(t)| \overset{P}{\to} 0 \tag{2.4}$$

as $n \to \infty$. Then by Gill (1980, Theorems 4.1.1 and 4.2.2) the Kaplan-Meier estimator (2.2) is uniformly consistent, and $\sqrt{n}(\hat{S}_n-S)$ converges weakly in the space $D[0,T]$ to $- S\cdot U$. Here U is a mean zero Gaussian process with $\mathrm{Cov}\{U(s),U(t)\} = \sigma^2(s\wedge t)$, where

$$\sigma^2(t) = \int_0^t \{\alpha(s)/y(s)\}ds \ . \tag{2.5}$$

Moreover n times $\hat{\sigma}_n^2(t)$, defined by (2.3), is a uniformly consistent estimator for $\sigma^2(t)$ given by (2.5).

The standard $100(1-\alpha)$ per cent confidence interval for $S(t)$ for a fixed $t\in[0,T]$ is

$$\hat{S}_n(t) \pm c_{\alpha/2}\, \hat{S}_n(t)\hat{\sigma}_n(t) \ , \tag{2.6}$$

where $c_{\alpha/2}$ is the upper $\alpha/2$ fractile of the standard normal distribution. To derive confidence intervals with better small sample properties we will consider the transformations $g(x) = \log(-\log x)$ and $g(x) = \arcsin \sqrt{x}$, the latter being variance stabilizing for the situation with no censoring. The log-minus-log transformation gives the $100(1-\alpha)$ per cent interval

$$\hat{S}_n(t)^{\exp\{\pm c_{\alpha/2}\hat{\sigma}_n(t)/\log \hat{S}_n(t)\}} \tag{2.7}$$

while the arcsine-square root transformation gives the interval

$$\begin{aligned} &\sin^2\{\max[0, \arcsin(\hat{S}_n(t)^{\frac{1}{2}}) - \tfrac{1}{2}c_{\alpha/2}\ \hat{\sigma}_n(t)(\hat{S}_n(t)/(1-\hat{S}_n(t)))^{\frac{1}{2}}]\} \\ &\leq S(t) \leq \\ &\sin^2\{\min[\pi/2, \arcsin(\hat{S}_n(t)^{\frac{1}{2}}) + \tfrac{1}{2}c_{\alpha/2}\ \hat{\sigma}_n(s)(\hat{S}_n(t)/(1-\hat{S}_n(t)))^{\frac{1}{2}}]\} \ . \end{aligned} \tag{2.8}$$

These will for short be denoted the logarithmic- and the arcsine-transformed confidence intervals.

We then consider simultaneous confidence bands for S on $[t_1, t_2]$, where $0 \leq t_1 < t_2 \leq T$. We introduce

$$c_i = \sigma^2(t_i)/\{1+\sigma^2(t_i)\} \tag{2.9}$$

and

$$\hat{c}_i = n\hat{\sigma}_n^2(t_i)/\{1+n\hat{\sigma}_n^2(t_i)\} \tag{2.10}$$

for $i=1,2$. Then the $100(1-\alpha)$ per cent confidence band derived by Hall & Wellner (1980) is given as

$$\hat{S}_n(t) \pm n^{-\frac{1}{2}} e_\alpha(\hat{c}_1, \hat{c}_2)(1+n\hat{\sigma}_n^2(t))\hat{S}_n(t) \ , \tag{2.11}$$

where $e_\alpha(c_1, c_2)$ is the upper α fractile in the distribution of $\sup_{c_1 \leq x \leq c_2} |W^0(x)|$ with W^0 being the standard Brownian bridge. It reduces to the well-known Kolmogorov band for completely observed survival data. For this band one will typically let $[t_1, t_2]$ be the whole of $[0, T]$, in which case tables of $e_\alpha(c_1, c_2) = e_\alpha(0, c_2)$ are given by Koziol & Byar (1975) and Hall & Wellner (1980). We will denote (2.11) the (non-transformed) Hall-Wellner band or HW band for short.

Nair (1984) derived the (non-transformed) equal precision band or EP band. This $100(1-\alpha)$ per cent confidence band is proportional to the pointwise one (2.6) and is given by

$$\hat{S}_n(t) \pm d_\alpha(\hat{c}_1,\hat{c}_2)\hat{S}_n(t)\hat{\sigma}_n(t) \,, \tag{2.12}$$

where $d_\alpha(c_1,c_2)$ is the upper α fractile in the distribution of $\sup_{c_1<x<c_2} |W^0(x)\{x(1-x)\}^{-\frac{1}{2}}|$. This fractile may be found by the asymptotic approximation of Miller & Siegmund (1982, formula (8)). The EP band is valid on intervals $[t_1,t_2]$ with $0<c_1<c_2<1$, cf. (2.9).

Using the argument of Bie, Borgan and Liestøl (1987, Section 2) and the asymptotic properties of the Kaplan-Meier estimator reviewed above, it is seen that the transformations $g(x) = \log(-\log x)$ and $g(x)=\arcsin\sqrt{x}$ yield bands similar to (2.7) and (2.8) respectively. We get bands proportional to the pointwise ones by replacing $c_{\alpha/2}$ by $d_\alpha(\hat{c}_1,\hat{c}_2)$ in (2.7) and (2.8). These bands, which we will denote the logarithmic- and arcsine-transformed EP bands, are valid on $[t_1,t_2]$, where $0<t_1<t_2<T$ are such that $0<c_1<c_2<1$; cf. (2.9). If we instead replace $c_{\alpha/2}\hat{\sigma}_n(t)$ by $n^{-\frac{1}{2}}e_\alpha(\hat{c}_1,\hat{c}_2)(1+n\hat{\sigma}_n^2(t))$ in (2.7) and (2.8) the logarithmic- and arcsine-transformed HW-bands result. These bands are valid on $[t_1,t_2]$, where $0<t_1<t_2<T$ with $S(t_1)<1$.

3. Survival and censoring distribution, simulation technique

The simulations were performed as described in Bie et al. (1987, Section 4). Using the random censorship model, various survival distributions and two censoring distributions were simulated. One of the survival distributions was the exponential one with parameter 1, the other two were the Weibull (1.35,2) and Weibull ($\sqrt{2}$,0.5) distributions having survival functions $e^{-1.35\, t^2}$ and $e^{-\sqrt{2}\, t^{0.5}}$, respectively. The censoring distributions were either exponential or uniform, with parameter adjusted to obtain the desired degree of censoring (usually 50%). The simulation results presented in the tables are based on 10 000 replications, corresponding to a standard error of the estimated level of confidence $(1-\alpha)$ or error rate (α) of about 0.002.

4. Pointwise confidence intervals. Small sample properties

Table 1 shows the error rates of confidence intervals with nominal confidence level 95% at 3 points in time, for two combinations of survival and censoring distributions and for n equal to 25, 50 and 200. As earlier shown by Thomas and Grunkemeier (1975), the error rates obtained when applying the standard interval (2.6) are too high, especially for $n=25$. An improvement is obtained by applying one of the transformed intervals. Although the logarithmic-transformed interval (2.7) gives slightly too low error rates, and the arcsine-transformed interval (2.8) slightly too high rates, the achieved confidence levels are quite acceptable even for $n=25$.

Similar results were obtained with the other combinations of survival and censoring distributions. Exceptions from this occur in situations where very few deaths are expected or where the expected number of individuals still at risk is very low. One such example is seen in Table 1 for the Weibull (1.35,2) distribution at $t=0.2$, where the expected number of deaths is only slightly above 1.

Thomas & Grunkemeier (1975) pointed out that the standard interval falls above and below the true parameter value an unbalanced number of times. Table 2 shows that this also occur for the transformed intervals. The arcsine-transformed interval seems, however, to be clearly better than the other two intervals in this respect.

With a nominal level of confidence of 99%, the approximation generally becomes less satisfactory. For example, when both the survival and censoring distributions are standard exponential and $t=0.4$, the standard interval (2.6) achieves the error rates (based on 20 000 simulations) 0.034 for $n=25$, 0.018 for $n=50$ and 0.010 for $n=200$. The corresponding numbers for the logarithmic-transformed interval (2.7) are 0.010, 0.011 and 0.009, while for the arcsine-transformed interval (2.8) they are 0.016, 0.012 and 0.009. This implies that a larger number of individuals is needed to achieve acceptable confidence levels when the nominal level is 99%. It seems as if roughly a doubling of the number of individuals is needed to obtain the same relative precision for a 99% interval as for a 95% interval.

An increase in the amount of censoring will as expected decrease the performance of the confidence intervals. However, even with 75% censoring, the transformed intervals produce acceptable results. For example, for the case where both survival and censoring distributions are exponential and $t=0.4$, the 95% standard interval (2.6) achieve the error rates 0.12 for $n=25$, 0.08 for $n=50$ and 0.06 for $n=200$. The corresponding numbers for the logarithmic-transformed interval (2.7) are 0.05, 0.05 and 0.05, while for the arcsine-transformed interval they are 0.08, 0.06, 0.05.

A comparison of the achieved error rates with the expected number of deaths in our simulations, indicate that with an expected number of deaths of at least 5, and a similar number still at risk, the arcsine transformed interval will have error rates around or below 0.06, while the logarithmic-transformed interval will have error rates above 0.04, the nominal error rate being 0.05. About 15-20 expected deaths seem necesssary to achieve the same precision with the standard interval. To obtain a balanced number of failures above and below the true value of the survival probability a higher number of deaths is necessary. However, the arcsine-transformed interval seems to produce reasonably balanced results even with 5-10 expected deaths.

5. Simultaneous confidence bands. Small sample properties

Table 3 shows the achieved error rates for the EP-band, the HW-band and their transformed versions when both survival and censoring distributions are standard exponential. As shown by Nair (1984), the non-transformed EP-band gives too high error rates when the number of observations is low. The table also shows that the improvement obtained by applying one of the transformed EP-bands is substantial. For the HW-band, the error rates are close to the nominal value, and the improvement obtained by using one of the transformations is of less importance.

An examination of the number of times the bands are located above and below the correct value at different points in time (results not shown) reveals that for the EP-bands, the arcsine transformed version performes clearly better than the other two. For the HW-bands, however, the improvement by using one of the transformed versions seems to be small also with respect to the error pattern.

6. Concluding remarks

The present study shows that a fairly low number of individuals is necessary to obtain reasonably reliable confidence intervals and bands for the survival function. For the intervals and the EP-band substantial improvements can be obtained by the use of one of the transformations. For the HW-band, there seems to be less reason for using one of the transformed versions.

Considering the confidence intervals, our opinion is that about 10 deaths, and a similar number still at risk, are sufficient for the transformed intervals to perform satisfactory, and even with 5 individuals in these categories, the intervals will probably be accurate enough for many purposes. The logarithmic-transformed interval tends to be slightly conservative and achieves confidence levels which are closer to the nominal ones than those obtained by the arcsine-transformed interval. On the other hand the arcsine-transformed interval gives a more symetric distribution of the failures above and below the true value of the survival probability.

Based on a restricted product limit estimator, Thomas & Grunkemeier (1975) gave two alternatives to the standard confidence interval. Due to different designs of the simulations, it is difficult to compare the small sample properties of these intervals and ours, but their performance seems to be comparable to that of the arcsine-transformed interval. Thomas & Grunkemeier's intervals do, however, require iterative computations and their idea is difficult to generalize to confidence bands.

It is difficult to state the necessary sample size for confidence bands, but if only a rough estimate of the uncertainty of the estimated survival curve is wanted, a band based on 20 (or even fewer) deaths seems to be satisfactory.

Although transformations are useful in confidence estimation of survival curves, the advantages are clearly smaller than those earlier reported for the cumulative hazard rate function (Bie et al, 1987).

Acknowledgement

We are grateful to Ole Bie for help in carrying out the simulations.

References

Aalen, O.O. & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scand. J. Statist. 5, 141-150.

Bie, O., Borgan, Ø. & Liestøl, K. (1987). Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties Scand. J. Statist. 14, 221-233.

Breslow, N. & Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. 2, 437-453.

Csörgő, S & Horváth, L. (1986). Confidence bands from censored samples. Cand. J. Statist. 14, 131-144.

Gill, R.D. (1980). Censoring and stochastic integrals. Math. Centre Tracts 124, Mathematical Centre, Amsterdam.

Gillespie, M.J. & Fisher, L. (1979). Confidence bands for the Kaplan-Meier survival curve estimate. Ann. Statist. 7, 920-924.

Hall, W.J. & Wellner, J.A. (1980). Confidence bands for a survival curve from censored data. Biometrika 67, 133-143.

Kalbfleisch, J.D. & Prentice, R.L. (1980). The statistical analysis of failure time data. Wiley, New York.

Kaplan, E.L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

Miller, R. & Siegmund, D. (1982). Maximally selected chi-square statistics. Biometrics 38, 1011-1016.

Nair, V.N. (1981). Plots and tests for goodness of fit with randomly censored data. Biometrika 68, 99-103.

Nair, V.N. (1984). Confidence bands for survival functions with censored data: A comparative study. Technometrics 26, 265-275.

Thomas, D.R. & Grunkemeier, G.L. (1975). Confidence interval estimation of survival probabilities for censored data. J. Amer. Statist. Assoc. 70, 865-871.

Table 1. Achieved error rates of confidence intervals with nominal level of confidence 95%

Survival/ censoring distributions		Standard interval (2.6)			Logarithmic-transformed interval (2.7)			Arc sine-transformed interval (2.8)		
	t=	0.2	0.6	1.0	0.2	0.6	1.0	0.2	0.6	1.0
Exponential/	n=25	0.07	0.08	0.09	0.05	0.05	0.04	0.07	0.06	0.06
exponential[2]	n=50	0.08	0.07	0.07	0.04	0.05	0.05	0.06	0.06	0.06
	n=200	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05
Weibull/	n=25	0.31	0.08	0.07	0.34	0.05	0.03	0.32	0.06	0.04
exponential[1]	n=50	0.10	0.07	0.08	0.13	0.05	0.04	0.11	0.06	0.06
	n=200	0.07	0.05	0.05	0.04	0.05	0.05	0.06	0.05	0.05

1 Both survival and censoring distributions are standard exponential.

2 Survival distribution is Weibull (1.35,2), censoring distribution is standard exponential.

Table 2. The distribution of the errors above and below the true value of the survival function at t=0.4 for confidence intervals with nominal level 95%. Survival and censoring distributions are both standard exponential.

	Standard interval (2.6)		Logarithmic-transformed interval (2.7)		Arc sine-transformed interval (2.8)	
	Above	Below	Above	Below	Above	Below
n=25	0.056	0.018	0.009	0.040	0.034	0.022
n=50	0.040	0.022	0.014	0.036	0.029	0.025
n=200	0.030	0.022	0.020	0.029	0.026	0.024

Table 3. Achieved error rates of confidence bands with nominal level of confidence 95%. Survival and censoring distributions are both standard exponential.

	EP-band (2.12)[1]	HW-band (2.11)	Logarithmic-transformed EP-band[1]	Logarithmic-transformed HW-band	Arc sine-transformed EP-band[1]	Arc sine-transformed HW-band
n=25	0.11	0.07	0.07	0.06	0.04	0.05
n=50	0.09	0.06	0.06	0.06	0.05	0.06
n=100	0.07	0.05	0.06	0.06	0.05	0.06
n=200	0.07	0.05	0.06	0.05	0.05	0.05

[1]) The bands are evaluated with $\hat{c}_1 = 0.05$ and $\hat{c}_2 = 0.95$.